DOCUMENT RESUME

ED 249 267 TM 840 620

AUTHOR Poggio, John P.
TITLE Practical Considerations When Setting Test Standards: A Look at the Process Used in Kansas.
PUB DATE Apr 84
NOTE 13p.; Paper presented at the Annual Meeting of the American Educational Research Association (68th, New Orleans, LA, April 23-27, 1984).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Descriptive (141) -- Tests/Evaluation Instruments (160)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Criterion Referenced Tests; *Cutting Scores; Elementary Secondary Education; *Minimum Competency Testing; *Scoring; *State Legislation; State Programs; State Standards; Testing Programs
IDENTIFIERS *Angoff Methods; Ebel Method; Kansas; *Standard Setting

ABSTRACT

Kansas law requires setting passing scores for the reading and mathematics competency test for 2nd, 4th, 6th, 8th, and 11th grade students, administered annually since April of 1980. New objective-referenced tests are prepared each year. Both judgmental (Angoff, Ebel, and Nedelsky) methods and empirical (contrasting groups and borderline) methods have been used to set test standards. While no one method appears to identify true cut scores and cut score comparison over methods is consistent with other research, only the Angoff and Ebel methods are currently being used. While problems were found with all methods, empirical and Nedelsky methods were more confusing to participants and yielded lower standards. A survey approach has replaced panel judgment for data collection. It is more efficient, permits a broader input base, and produces more psychometrically favorable standards. A 26-member State Advising Committee inperpolates for the data gathered to set standards rather than using the mathematics prescribed by the methods. The process, while objective to a point, remains largely value-laden. Standard data for each 1982 test are given. Sample survey forms and rating sheets for the Angoff and Ebel methods are appended. (BS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. Poggio

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED249267

Practical Considerations When Setting Test Standards: A Look at the Process Used in Kansas

John P. Poggio
University of Kansas

Kansas Law requires competency testing of students at grades 2, 4, 6, 8 and 11. Students are tested in the areas of reading and mathematics. Testing has been carried out in April of 1980, 1982 and 1983. In each year, new objective-referenced tests are prepared. While legislation does not call for promotion decisions or diploma awarding to be tied to a student's performance on the tests, legislation does require that a passing score be set for each of the 10 tests. Public reports of building, district and state performance are oriented around the proportion of students, by area and grade level, judged minimally competent. As such, matters surrounding how the passing score was arrived at receive considerable attention. This paper describes how standards are set in the State, and highlights issues with which we have had to contend.

Methods Tried

Over the years we have attempted setting the test standards using judgmental (Angoff, Ebel and Nedelsky) methods and empirical (contrasting groups and boarder line) methods. We have used the judgmental methods in two formats. One format was to convene panels of judges and derive standards in the mode commonly cited and discussed. The second format has been to prepare a survey type questionnaire (see attachment) and send it to large samples of judges across the state for rating.

In the sections which follow are discussed what we have learned over time about each of the methods. It should be noted throughout this presentation that (1) no one method has surfaced as the one to use that

TM 840 620

identifies the true cut-score, and (2) comparisons of the cut-scores over methods is altogether consistent with other research represented in this area (Poggio and Glasnapp, 1981; 1982).

-- Empirical Methods: Contrasting Groups and Border Line

(1) the methods are rather easily implementable;

(2) teachers report little difficulty in following what is to be done in Contrasting Groups;

(3) the Boarder Line method does create some confusion with teachers, i.e., "What do you mean just barely minimally competent?"

(4) the standards that result tend to be lower than those yielded by Angoff or Ebel;

(5) the standard becomes available typically well after actual testing;

(6) the public is both confused and tends to doubt the legitimacy of the standard when they (often) cannot understand the "statistical magic" which delivers the standard; and

(7) the methods give support to the often stated contention that "teachers can already tell us who is competent."

— Judgmental Method: Nedelsky

(1) Judges find this method very confusing and they report not being confident in their judgments;

(2) it can only be used by experienced teachers;

(3) in implementing the method judges tend not to be careful in their study of items, often marking the correct choice as being not a viable distractor; and

(4) it delivers a standard substantially below that of all other methods. In this context data from it is quickly ignored.

— Judgmental Method: Angoff

(1) easy to implement and understand either in a panel format or in a survey format;

(2) judges tend to establish their own "mean" level, causing considerable variability among the individual judge standards. This becomes particularly problematic in the panel approach when so few judges are used;

(3) having to define students "who are minimally competent" is a problem for many judges;

—Judgmental Method: Ebel

(1) the task itself is time consuming. Fatigue and boredom can become a problem;

(2) the method is rather easy for most judges to understand and can be implemented without difficulty;

(3) the relevance rating position of "Questionable" causes judges to become concerned about the method;

(4) the cell percent passing task causes real difficulty/debate over the "Questionable" dimension; and

(5) when computing the standard, it can vary considerably depending whether it is computed by judge, or based on the group cell values.

In practice, we no longer use the Nedelsky or empirical methods for reasons given above. We rely on gathering standards data on <u>both</u> the Ebel and Angoff methods. Also we have abandoned the use of panels. We have found the survey approach: (1) to be far more efficient relative to time and cost, (2) the survey approach permits a broader base for input to the decision-making process, and (3) standards across the two approaches are comparable, and psychometrically favor the survey approach.

Yet, once the data are obtained the actual setting of each test's standard is not solved by the mathematics prescribed by the method. In fact, it is interpolated for the data gathered (see Table I) by a 26-member State Advising Committee.

The process, while objective to a point, remains largely value ladened.

References

Poggio, J. P., Glasnapp, D. R. and Eros, D. S. An empirical investigation of the Angoff, Ebel and Nedelsky standard setting methods. Papers presented at the annual meeting of the American Educational Research Association, Los Angeles, 1981.

Poggio, J. P., Glasnapp, D. R. and Eros, D. S. An evaluation of contrasting groups methods for setting test standards. Paper presented at the annual meeting of the American Educational Research Association, New York, 1982.

Poggio, J. P., Glasnapp, D. R. and Eros, D. S. An analysis of the validity of judgmental methods used to set test standards. Paper prevented at the annual meeting of the American Educational Research Association, Montreal, 1983.

Table I

STANDARD DATA - KANSAS MCT -1982

Grade	Area	Method	N	P_{40}	Mean	Median	P_{60}	Test Standard Used
2	Read.	Angoff	41	34.50	35.93	37.63	38.20	35
		Ebel	41	34.00	34.43	34.21	34.65	
2	Math.	Angoff	47	32.00	33.45	35.00	36.50	34
		Ebel	41	33.40	34.25	34.05	35.20	
4	Read.	Angoff	38	40.30	41.84	45.83	46.50	45
		Ebel	39	44.41	45.17	45.81	46.15	
4	Math.	Angoff	41	43.50	45.29	46.33	48.75	43
		Ebel	43	41.27	41.89	41.86	42.79	
6	Read.	Angoff	43	41.80	43.95	45.00	46.00	45
		Ebel	37	45.27	46.01	45.41	46.42	
6	Math.	Angoff	38	42.84	44.10	46.75	47.20	45
		Ebel	39	45.75	46.13	46.13	46.65	
8	Read.	Angoff	38	40.50	41.21	44.83	46.49	43
		Ebel	40	43.21	43.16	43.72	43.97	
8	Math.	Angoff	40	40.10	42.40	42.50	44.09	42
		Ebel	37	41.15	41.85	41.80	42.26	
11	Read.	Angoff	38	44.25	45.66	46.90	47.80	46
		Ebel	38	46.70	47.27	47.65	47.75	
11	Math.	Angoff	38	38.50	39.58	41.50	43.50	41
		Ebel	33	41.20	42.44	42.12	42.55	

SAMPLE SURVEY FORM USED WITH THE ANGOFF METHOD

DIRECTIONS FOR STANDARD SETTING

Attached you will find a copy of the Kansas Minimum Competency Test in Reading for Grade 11. Your task is to read an item, then estimate the probability that the minimally competent student in Grade 11 in Reading would answer the item correctly. You are to assign probabilities on a scale from 0 to 100, where 0 would mean that the minimally competent student has no chance of answering the item correctly, and 100 would mean that the minimally competent child is certain to answer the item correctly. Your estimate of the probability that an item will be answered correctly is to be recorded on the separate response sheet.

For example, consider the item:

(Sample) Which of the following is the opposite of happy?

A. glad
B. cry
C. jump
D. sad

If you believe that a minimally competent child has a 96 percent chance of answering this item correctly, then you would write 96 on the response sheet. You are to indicate the probability of a correct response by the minimally competent child to each item on the attached response sheet.

When making your judgements about each item, use the following guidelines:

1. Use your own definition of a minimally competent student at the grade level for the test you are reviewing. It is best to think of the skill level of a minimally competent group of students rather than a single individual.

2. Do <u>not</u> review the performance of your students on the test prior to making your judgements. Rather, let your expert opinion and experience dictate the likelihood of a correct response by a minimally competent student.

Use this sheet to rate: GRADE 11 READING

RESPONSE SHEET

DIRECTIONS: Read a test item, then judge the probability that the minimally competent student in this grade level for this content area would answer the item correctly. Probabilities can be assigned from 0 (no chance) to 100 (absolutely certain). Write the probability on the line next to each item number.

Item	Probability of Correct Answer	Item	Probability of Correct Answer	Item	Probability of Correct Answer
1	_____	21	_____	41	_____
2	_____	22	_____	42	_____
3	_____	23	_____	43	_____
4	_____	24	_____	44	_____
5	_____	25	_____	45	_____
6	_____	26	_____	46	_____
7	_____	27	_____	47	_____
8	_____	28	_____	48	_____
9	_____	29	_____	49	_____
10	_____	30	_____	50	_____
11	_____	31	_____	51	_____
12	_____	32	_____	52	_____
13	_____	33	_____	53	_____
14	_____	34	_____	54	_____
15	_____	35	_____	55	_____
16	_____	36	_____	56	_____
17	_____	37	_____	57	_____
18	_____	38	_____	58	_____
19	_____	39	_____	59	_____
20	_____	40	_____	60	_____

Thank you for your assistance with this activity. Please return these materials to the person who gave them to you.

SAMPLE SURVEY FORM USED WITH THE EBEL METHOD

ACTIVITY 1: DIRECTIONS FOR STANDARD SETTING

Attached you will find a copy of the Kansas Minimum Competency Test in Mathematics for Grade 8. Your task is to read an item, then make two separate judgements about the item, one for difficulty and one for relevance. Directions for making the judgements follow.

RATING FOR DIFFICULTY. After reading an item, judge how difficult the item is for Grade 8 students. Each item is to be rated as either:

EASY (E)
MEDIUM (M)
or HARD (H).

For example, consider the item:

(Sample) Which number is greater than 89?

A. 95
B. 80
C. 50
D. 10

If you believe this item is easy for Grade 8 students you would circle the E by this item on the response sheet. Difficulty ratings are to be made for each item and recorded on the response sheet before beginning the next rating task.

RATING FOR RELEVANCE. Each item tests a skill in mathematics. Next, after reading an item, judge how relevant the item is as a measure of a minimum competency skill in Grade 8. The item may be rated as either:

ESSENTIAL (E)
IMPORTANT (I)
ACCEPTABLE (A)
or QUESTIONABLE (Q)

For example, if you believe that the sample item is essential as a minimal competency for Grade 8 students in mathematics, you would circle E on the response sheet. Relevance ratings are to be completed for all items and recorded on the appropriate response sheet before beginning the final activity.

Use this sheet to rate: GRADE 8 MATHEMATICS

ACTIVITY 1: ITEM DIFFICULTY-RESPONSE SHEET

DIRECTIONS: Read each test item, then judge how difficult the item is for students at Grade 8. Rate each item on difficulty by circling:

E for easy
M for medium
H for hard

Item	Difficulty	Item	Difficulty	Item	Difficulty
1	E M H	21	E M H	41	E M H
2	E M H	22	E M H	42	E M H
3	E M H	23	E M H	43	E M H
4	E M H	24	E M H	44	E M H
5	E M H	25	E M H	45	E M H
6	E M H	26	E M H	46	E M H
7	E M H	27	E M H	47	E M H
8	E M H	28	E M H	48	E M H
9	E M H	29	E M H	49	E M H
10	E M H	30	E M H	50	E M H
11	E M H	31	E M H	51	E M H
12	E M H	32	E M H	52	E M H
13	E M H	33	E M H	53	E M H
14	E M H	34	E M H	54	E M H
15	E M H	35	E M H	55	E M H
16	E M H	36	E M H	56	E M H
17	E M H	37	E M H	57	E M H
18	E M H	38	E M H	58	E M H
19	E M H	39	E M H	59	E M H
20	E M H	40	E M H	60	E M H

Use this sheet to rate: GRADE 8 MATHEMATICS

ACTIVITY 1: ITEM RELEVANCE-RESPONSE SHEET

DIRECTIONS: Read each test item, then judge how relevant the item is as a measure of a minimal competency for students in Grade 8. Rate each item on relevance by circling:

E for essential
I for important
A for acceptable
Q for questionable

Item	Relevance	Item	Relevance	Item	Relevance
1	E I A Q	21	E I A Q	41	E I A Q
2	E I A Q	22	E I A Q	42	E I A Q
3	E I A Q	23	E I A Q	43	E I A Q
4	E I A Q	24	E I A Q	44	E I A Q
5	E I A Q	25	E I A Q	45	E I A Q
6	E I A Q	26	E I A Q	46	E I A Q
7	E I A Q	27	E I A Q	47	E I A Q
8	E I A Q	28	E I A Q	48	E I A Q
9	E I A Q	29	E I A Q	49	E I A Q
10	E I A Q	30	E I A Q	50	E I A Q
11	E I A Q	31	E I A Q	51	E I A Q
12	E I A Q	32	E I A Q	52	E I A Q
13	E I A Q	33	E I A Q	53	E I A Q
14	E I A Q	34	E I A Q	54	E I A Q
15	E I A Q	35	E I A Q	55	E I A Q
16	E I A Q	36	E I A Q	56	E I A Q
17	E I A Q	37	E I A Q	57	E I A Q
18	E I A Q	38	E I A Q	58	E I A Q
19	E I A Q	39	E I A Q	59	E I A Q
20	E I A Q	40	E I A Q	60	E I A Q

ACTIVITY 2: REQUIRED PERFORMANCE LEVELS

This activity involves making judgements about general categories of items. Based on difficulty level and relevance of items, 12 separate categories of items may be found on a test.

1. Consider a set of 100 test items all of which have been judged ESSENTIAL and HARD. How many of these 100 items should a student be able to answer correctly in order to be judged minimally competent?

   _____items

2. Consider a set of 100 test items all of which have been judged ESSENTIAL and of MEDIUM DIFFICULTY. How many of these 100 items should a student be able to answer correctly in order to be judged minimally competent?

   _____items

3. Consider a set of 100 test items all of which have been judged ESSENTIAL and EASY. How many of these 100 items should a student be able to answer correctly in order to be judged minimally competent?

   _____items

4. Consider a set of 100 test items all of which have been judged IMPORTANT and HARD. How many of these 100 items should a student be able to answer correctly in order to be judged minimally competent?

   _____items

5. Consider a set of 100 test items all of which have been judged IMPORTANT and of MEDIUM DIFFICULTY. How many of these 100 items should a student be able to answer correctly in order to be judged minimally competent?

   _____items

6. Consider a set of 100 test items all of which have been judged IMPORTANT and EASY. How many of these 100 items should a student be able to answer correctly in order to be judged minimally competent?

   _____items

7. Consider a set of 100 test items all of which have been judged ACCEPTABLE and HARD. How many of these 100 items should a student be able to answer correctly in order to be judged minimally competent?

_____items

8. Consider a set of 100 test items all of which have been judged ACCEPTABLE and of MEDIUM DIFFICULTY. How many of these 100 items should a student be able to answer correctly in order to be judged minimally competent?

_____items

9. Consider a set of 100 test items all of which have been judged ACCEPTABLE and EASY. How many of these 100 items should a student be able to answer correctly in order to be judged minimally competent?

_____items

10. Consider a set of 100 test items all of which have been judged of QUESTIONABLE RELEVANCE and HARD. How many of these 100 items should a student be able to answer correctly in order to be judged minimally competent?

_____items

11. Consider a set of 100 test items all of which have been judged of QUESTIONABLE RELEVANCE and of MEDIUM DIFFICULTY. How many of these 100 items should a student be able to answer correctly in order to be judged minimally competent?

_____items

12. Consider a set of 100 test items all of which have been judged of QUESTIONABLE RELEVANCE and EASY. How many of these 100 items should a student be able to answer correctly in order to be judged minimally competent?

_____items

Thank you for your assistance with this activity. Please return these materials to the person who gave them to you.